HUMAN-MACHINE INTERACTION WITH MULTIPLE AUTONOMOUS SENSORS

Steven A. Murray

Navy Command, Control and Ocean Surveillance Center, RDT&E Division,
San Diego, California

Security and inspection systems are becoming increasingly automated. Many such systems include mobile platforms capable of autonomous sensing and analysis of the environment from a multitude of perspectives. This increased automation shifts the responsibilities of humans from active patrolling and inspection to passive monitoring of remote sensor information. The operator brings perceptual and cognitive characteristics to this task which need to be addressed in both system architecture and interface design if desired performance reliability is to be achieved. A study is reported which examines the impact of these characteristics on system performance.

Keywords: Human-centered design, human-machine interface, human reliability.

1. INTRODUCTION

The use of autonomous systems for remote sensing and inspection has long been an important design effort of system engineering. Autonomous sensing can preclude the need for humans to physically patrol large areas, such as factories or warehouses, and can protect them from hazardous environments. In addition, such systems do not fatigue, or vary significantly in detection performance over time, as humans often do. Autonomous systems, however, are not foolproof. Sensors can fail to detect important events, or can report false alarms as a function of physical fault or imperfect analysis algorithms. Human monitoring is therefore included in most system designs to ensure proper operation and improved signal classification (e.g., Everett *et al*., 1992). Furthermore, to realize the best economic potential of such systems, a single operator is often responsible for supervising several remote platforms simultaneously. This leveraging of human presence presumes that significant events will occur for only a fraction of the platforms at any one time; that is, designers assume that "worst case" events will still be within the response capabilities of the human operator.

Controlling, or even monitoring multiple-platform systems such as these can complex if the areas under surveillance are large or if many sensors are employed. Because human performance tends to degrade over time when monitoring systems with low event rates (the "vigilance decrement;" Parasuraman, 1986), such automated systems usually provide for operator cueing when a significant event occurs by delivering an alert signal and/or information about the event. This latter case can involve presentation of a video image covering the area where the detection was first triggered. The operator must confirm the nature of the event, locate it in physical space (e.g., Is it inside or outside the building? Which hallway or room is it in? etc.), and initiate an appropriate response. Human-machine interfaces for this task almost invariably provide an overall, two-dimensional depiction of the surveilled area -- such as a building diagram, displayed on a CRT -- so that the location of the alerting platform can be determined and the video image correlated with a point in physical space. Thus, the operator must rapidly map information between 2-D to 3-D representations of the environment. In addition, many sensors have panning capability, so the location of a sensor platform does not fully define the location of its image. Although some systems provide field-of-view markers to aid this task, the 2-D to 3-D information mapping must still be performed.

Operators can learn, over time, what to expect from sensor images of conventional security and monitoring systems (e.g., those with fixed, wall-mounted cameras), based on their known location and orientation in the surveilled area. The use of mobile sensor platforms, however, considerably complicates the operator's control task by removing the predictability of fixed sensor images. Each alert must be independently interpreted, based on the current location of the reporting platform(s). Furthermore, independent patrol patterns of autonomous platforms open the possibility of multiple signals arising from the same physical event due to overlapping fields of coverage (i.e., during those occasions when patrol patterns come into close proximity). Although such redundancy may complicate the operator's task, there may be sound reasons for desiring such overlaps in sensor coverage, such as increased detection reliability. Therefore, redundant reports may be expected to occur in some fraction of system events. In these situations, the operator must map each image in space, based on platform location, and must correlate information across images, to generate an integrated spatial model of the events being reported.

Development of multiple-platform, autonomous systems such as these, for both security and process monitoring, are underway at several government and commercial laboratories, including the Naval Command, Control, and Ocean Surveillance Center. These systems support both structured and unstructured patrolling by autonomous robots, which search for objects or events based on loosely-structured criteria (e.g., some combination of movement, contrast, or temperature thresholds). The wide variety of settings in which these systems need to function make stringent demands on automation capabilities; human presence is very much needed to ensure reliable performance and appropriate follow-up response. An information flow diagram is shown in Figure 1.

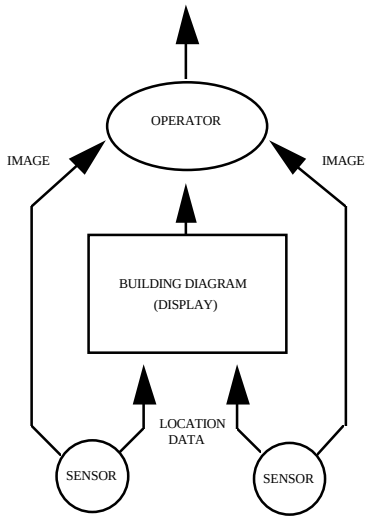

Figure 1. Processing and control diagram of multiple-sensor security system.

Depending on the degree of automation and the complexity of the task (e.g., building security with fixed cameras, or outdoor surveillance with mobile robots), the operator can spend most of the time in a passive role. All sensor control and raw data analyses are handled automatically, and the operator is required to interact with the system only when alerted. From this point, all subsequent performance is essentially manual, as the operator searches and interprets each display, takes the appropriate action (such as directing a response to the location of the event), and releases the sensors to resume surveillance. Control is, thus, more "traded" than "shared" (Sheridan, 1992, page 65). Models of human supervisory control (e.g., Rasmussen, 1983; Sheridan, 1987) are not readily applied to tasks such as these, where the operator is essentially decoupled from the system until an alert effectively forces a change to exclusively manual control. A problem with this approach, as in many automated system designs (Woods, in press) is that the operator is required to perform at precisely those moments when the system can provide no further assistance, i.e., the sensors have done their job of detecting and relaying events. These moments are also those conditions (e.g., suspected intrusions, fires, etc.) for which prompt and correct action is essential.

Performance estimates for such complex systems often rely on extrapolations of operator models generated from much simpler, single-sensor applications. Consequence of this approach may include an overly-optimistic assessment of what the human-machine system can accomplish, assignment of too many systems to the control of a single person, or inadequate human-machine interface design. Different interface modeling and design approaches may be required to support operator performance in such one-to-many control situations. An initial examination of human performance capabilities in such settings was therefore conducted to demonstrate the limitations of single-sensor models for predicting operator performance, and to identify certain task variables in need of special interface support.

2. PERFORMANCE STUDY

The selection of an industrial security system as the experimental setting was dictated by practical need; a research effort was required to support the development of prototype systems which used multiple autonomous robots as sensing and reporting platforms. The task nevertheless illustrates general human capabilities to rapidly assess (pictorial) sensor information from a number of distributed sources and to integrate it into a single model of the environment, which should have wider application.

Three task characteristics were selected for initial study: the number of displays which had to be monitored (corresponding to the number of remote sensor systems), the amount of information to be interpreted (i.e., the number of images which were presented at any one time), and the complexity of the information (i.e., whether the images depicted separate events or a common event from overlapping sensors).

2.1 Hypotheses

A number of displays used for this experiment was limited to three. Each display represented a potential source of task-relevant information from a remote sensor, i.e., the presence of an image indicated that a sensor had detected something significant, and the operator was therefore compelled to examine the image, if only to confirm a false alarm. Although not all images contained a target (for this application, a target was a simulated human "intruder"), all targets were clearly visible in those images where they were presented. Target figures were all replications of the same model, i.e., the figures all looked the same but were positioned at different locations within the environment. Because the major influence on search time is the size of the search set (Hyman, 1953; Scanlan, 1977), operator response time was predicted

to increase with larger numbers of displays and with greater numbers of target figures.

Task complexity was manipulated by controlling image redundancy. High redundancy was defined as multiple images of the same object, i.e., each display showing a different perspective of the same target figure. Low redundancy was defined as a separate object in each image. It was hypothesized that the effort involved in correlating similar scenes, to resolve the number of actual targets present, would be a more complex task than resolving images containing distinct targets and, therefore, high-redundancy conditions would require more processing time. This is also in accord with a prediction by Vickers (1970) of increased reaction time as a function of reduced inter-stimulus discriminability.

It was further hypothesized that additional cognitive processing demands would be required for this task if the operator had to map images from unpredictable platform locations (and viewpoints) for each trial, i.e., if the operator had to interpret images from mobile platforms. This additional workload should be evidenced by poorer performance for trials imaged from mobile sensor platforms than trials imaged from fixed-position platforms.

To make the task as realistic as possible, and to control for individual speed-accuracy tradeoff strategies (e.g., Wickens, 1984), subject instructions for this experiment emphasized accuracy; operators were told that response accuracy was more important than speed. Response times were therefore expected to be more informative than error rates as a dependent measure.

2.2 Method

Subjects. Six volunteers from the laboratory staff (four males and two females, aged 26 to 32) were used as subjects. All participants were familiar with the task environment used in the simulation.

Stimulus preparation. An indoor setting -- a single, open-bay building -- was selected as the task environment. This setting was already modeled on a Silicon Graphics computer system, and had been previously used for virtual environment applications. A set of static scenes was generated from this model by moving through it with a simulated sensor platform (about four feet above ground level) and capturing images at various locations and along various directional bearings. Images were monochromatic, and approximately 9.0 x 7.5 cm in size. A human figure was modeled in the simulation to serve as an "intruder." Copies of this figure were placed at varying locations in a subset of these images to serve as visual targets. No more than one target was contained in any given image. Images were displayed on a Silicon Graphics Indigo computer with a 43 cm (diagonal), high-resolution (1280 x 1024 pixel) monitor.

A paper diagram of the building was provided for each trial, which depicted major features of the interior (e.g., furniture, doors) and which contained the location of the sensor(s) which had generated the associated video image(s). An example of such an image - diagram pair is presented in Figure 2. The illustration shows an image (from a single sensor) of a figure standing at the end of a hallway, in front of a door. Subjects used these diagrams to record their task responses.

Design. A repeated measures design was used, which consisted of two conditions for Sensor Mobility (i.e., fixed-position platforms versus platforms which were free to change position between trials), three levels of Number of Displays (i.e., one, two, or three displays, presented

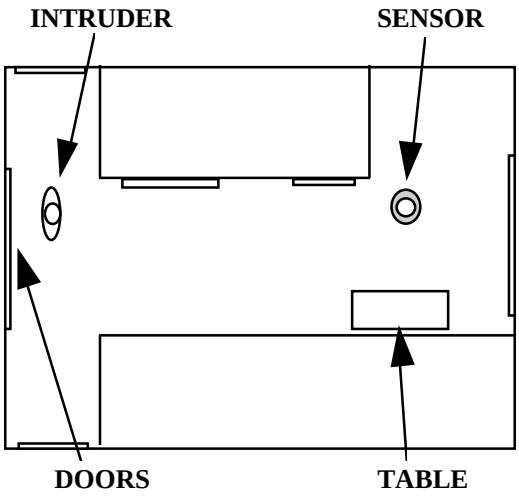

Figure 2. Simulated video image of intruder with diagram of corresponding object locations

simultaneously from a corresponding number of simulated sensor platforms), three levels of Number of Figures (i.e., how many images actually contained a target figure), and three levels of Redundancy (i.e., HIGH, where all displays showed a common target, MEDIUM, where some displays showed a common target, and LOW, where each display showed a unique target).

A decision was made that each image would contain, at most, a single target figure. This necessarily resulted in an incomplete design (i.e., some cells of the 2x3x3x3 factorial were not used). Target detection across displays was the performance of interest. If a full factorial design had been employed, some displays would have contained multiple targets, allowing the task to be completed by straightforward counting; the desire to avoid this confounding behavior led to the design approach described.

Procedure. The nature of the security monitor job was explained to the subjects using standardized verbal instructions. Response accuracy was emphasized by explaining that erroneous interpretations would result in time being wasted by response forces being dispatched to the wrong location. A set of training trials was then administered, which contained at least one example of every condition to be encountered in the actual experiment. A trial began when the images appeared on the monitor, and ended when the subject had marked the position(s) of the intruder(s) on the paper diagram; the experimenter then pressed a key which recorded elapsed time for the trial. Subjects were always told whether they were dealing with fixed or mobile platforms, and how many sensors were active in the environment. Other conditions were randomized. Every trial had at least one image with a target figure in it; i.e., there were no completely "false alarm" trials.

2.3 Results

Main effects from an analysis of variance were all significant and were in expected directions. Because the incomplete nature of the design cells prevented calculation of a complete ANOVA , a multiple regression analysis was also performed to gain an understanding of the simultaneous action of these variables.

ANOVA effects. The results of the sensor mobility manipulation showed that responses to images from mobile sensor platforms took approximately forty percent more time than responses to images from fixed-position sensors; $\underline{F}(1,5) = 8.97$, $\underline{p} < .001$. Response time also increased as a function of monitoring increasing numbers of displays ($\underline{F}(2,10) = 126.31$, $\underline{p} < .001$) and as a function of the number of target figures shown on those displays ($\underline{F}(2,10) = 95.69$, $\underline{p} < .001$). The redundancy manipulation showed that system operators took longer to process multiple independent images than they did to process redundant ones ($\underline{F}(2,10) = 77.11$, $\underline{p} < .001$). Mean response times for subjects are presented in Table 1.

Table 1. Mean response times results, by condition

response time(sec)			mean
Sensor Mobility	fixed		8.91
	mobile		12.15
Number of Displays	1		5.96
	2	9.89	
	3		11.61
Number of Figures	1		6.97
	2	10.59	
	3		14.01
Redundancy	high		8.53
	medium		12.45
	low		16.59

Multiple regression. A forward stepwise regression analysis was performed separately for the response time results of the fixed-platform and the mobile-platform conditions, because it is unlikely that these two sensor positioning schemes would ever exist in a single, hybrid system. Results showed that levels of image redundancy and the number of figures shown on the displays contributed the most to response time performance for both types of systems. The number of displays which the operator had to monitor did not account for enough additional variance to be included in either regression equation. For the fixed-platform condition:

Response Time (sec) = 2.440 + 0.478 (Redundancy) + 0.328 (Number of Figures) (1)

R^2 for this equation was 0.5157, $F(2,117) = 62.762$, $p < .001$. For the mobile-platform condition:

Response Time (sec) = 2.035 + 0.477 (Number of Figures) + 0.343 (Redundancy) (2)

R^2 for this equation is 0.5342, $F(2,117) = 67.083$, $p < .001$.

3. DISCUSSION

This experiment succeeded in highlighting the relative importance of selected system and display variables to operator performance, and provided some indication of the sensitivity of performance to manipulations of those variables.

3.1 Analysis

In keeping with the priorities established for the security monitor's job, response time proved to be an effective performance measure. Subjects appeared to trade response time for some threshold level of accuracy. It is possible, however, that an operator strategy that emphasized rapid response might show similar patterns of results for error rates (e.g., by identifying an incorrect number of intruders than was actually the case, or by locating them at incorrect positions in the building). Furthermore, operators in real world settings with larger numbers of displays might demonstrate even poorer performance than obtained in this experiment. A modern industrial security system, for example, may have as many as 125 displays under the control of a single operator. The most important consequence of such data is that they demonstrate a potential bottleneck on total system performance; the system can only be as fast and accurate as the operator, who is the final filter on input data and the only initiator of system action.

The most significant factor influencing task performance was the use of autonomous, mobile sensors, i.e., platforms whose physical location (and viewpoint) could change from one trial to the next. Clearly, information from such sensors took longer to interpret than information from fixed-position sensors. The flexibility and expanded sensor coverage afforded by using such an autonomous system therefore comes at a price in operator workload, and might.

The experiment showed shorter response times for trials with redundant images. For a given number of images, subjects were apparently able to determine commonality more rapidly than uniqueness. This was not a trivial task, however, as different viewpoints used for redundant images were made at different (apparent) distances from the target figures, as well as different viewing angles. These changes resulted in different effective fields of view, and thus to changes in both aspect and relative size of the target figures, and to shifts in the contents of scene backgrounds.

Subjects used in the study were familiar with the visual environment of the simulation, and knew both the building layout and the locations of its contents (e.g., doors, windows, shelves, etc.). This result, therefore, could have been a function of subject experience with the particular environment used in the experiment, or could have been obtained by some process that worked equally well with any set of redundant images. Although the issue cannot be completely resolved here, the results of the sensor mobility manipulation would seem to support an explanation based on the inherent redundancy of the images. If familiarity with the environment were the essential factor, this manipulation would probably not have shown such a large difference in performance, as all images -- fixed or mobile -- contained portions of the "familiar" background. This conclusion is indirectly supported by other research (Thorisson, 1993), that used both reaction time and eye movements to determine that subjects could extract three-dimensional information from multiple two-dimensional images using a feature search process; mental reconstruction of a scene in three-dimensions was not necessary.

3.2 Application

Results of this study provide some guidelines for predicting human-machine performance for systems involving multiple, autonomous sensors. As stated earlier, systems such as these already exist and more sophisticated versions are being developed. The rapid increase in response time for even the modest levels of manipulations used here is cause for concern, especially when newer systems are planned with larger numbers of sensors and are designed for operations in cluttered environments.

Operator activities in systems like these do not match the common functions typical of supervisory control (Sheridan, 1992, chapter 1). Certainly, the closed feedback loops found in many human-machine control systems are not present. Nevertheless, the operator has a central control function, as it is the operator who filters and transforms sensor products to produce the final system output. The human-machine interface, as the only transfer point of sensor output to the operator, is therefore essential to system performance. Woods (in press) has written extensively about the need for human-machine design which supports the extraction of task-relevant meaning from input data, rather than the mere delivery of data to the operator, i.e., to design for information extraction at the whole task level. Like the computer applications which Woods addresses, each image from a sensor platform acts as one "keyhole" into a much larger data space. It is the coordination of multiple views, or "keyholes," into a single picture of this space that is not supported by the design approaches examined here.

Additional measures could be exploited toward this end, to improve the human-machine interface and, thereby, to enhance system performance. Providing additional visual cues (such as directional lines from each sensor) on the diagram display, for example, could resolve sensor views by identifying overlaps where those lines crossed. This approach would still require operator processing, however, and consume additional response time, especially for systems using higher numbers of sensors. An alternative would be to provide an inhibitory feature to the display of such redundant events (Sheridan, 1992, page 289), whereby only one alert is provided regardless of the number of overlapping contacts. This would require additional computer processing, however, to automatically detect such redundancy. Both of this design concepts are readily testable, and additional investigations are being initiated to measure their effects.

A larger issue may be the consequence of accessing human intervention only when needed in an otherwise-automated setting. The operator is present to fill performance "holes" in the sensing and classification process. The event-driven nature of this approach, because events are sporadic (yet important), complicates any effort by the operator to maintain a current mental model of the environment. An alternative interface scheme is to reduce the complexity of the 3-D to 2-D image transformations when events are detected by providing a 2.5 or 3-D rendition of the layout of the surveilled area. The operator task would then involve smaller spatial transformations when examining sensor images, would assist in visualization of the entire suite of sensor patterns, and might help to keep the operator more consistently involved by expanding the utility a display which is always present (i.e., unlike sensor images). This, too, is being investigated.

System applications continue to emerge which require extensions to models of human-machine interaction, and which motivate empirical measurement to establish and scale critical variables. The use of multiple, autonomous sensor systems is one such application. Solutions to these problems can, in turn, enhance human-machine design for a variety of other engineering needs, as well.

ACKNOWLEDGMENTS

This research was supported by an Office of Naval Research contract N0001493WX24310AA, under the administration of Dr. Harold Hawkins.

REFERENCES

Everett, H.R., Gilbreath, G.A., and Laird, R.R. (1992). Multiple robot host architecture. *NRaD Technical Note 1710*.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *J. Exp. Psychol.*, **45**, 423-432.

Parasuraman (1986). Vigilance, monitoring, and search. In: *Handbook of Perception and Human Performance* (K.R. Boff, L. Kaufman, and J.P. Thomas, Eds.), Vol. 2, Chap. 43, pp. 1-39. John Wiley, New York.

Rasmussen, J. (1983). Skills, rules and knowledge: signals, signs, and symbols, and other distinctions in human performance models. *IEEE Trans. Systems, Man and Cybernetics*, **SMC-133**, 257-267.

Scanlan, L.A. (1977). Target acquisition in realistic terrain. In: *Proceedings, 21st Annual Meeting of the Human Factors Society* (A.S. Neal and R. Palasek Eds.), Human Factors Society, Santa Monica, CA.

Sheridan, T.B. (1987). Supervisory control. In: *Handbook of Human Factors* (G. Salvendy, Ed.), Chap. 9, pp. 1243-1268. John Wiley, New York.

Sheridan, T.B. (1992). *Telerobotics, Automation, and Human Supervisory Control*. MIT Press, Cambridge, MA.

Thorisson, K.R. (1993). Estimating three-dimensional space from multiple two-dimensional views. *Presence*, **2(1)**, 44-53.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, **13**, 37-58.

Wickens, C. (1984). *Engineering Psychology and Human Performance*. Charles E. Merrill, Columbus, OH.

Woods, D.D. (1988). Coping with complexity: The psychology of human behavior in complex systems. In: *Mental Models, Tasks and Errors* (L.P. Goodstein, H.B. Andersen, and S.E. Olsen Eds.). Taylor & Francis, London.

Woods, D.D. (in press). Towards a theoretical base for representation design in the computer medium: ecological perception and aiding human cognition. In: *The Ecology of Human-Machine Systems: A Global Perspective* (J. Flach, P. Hancock, J. Caird and K. Vicente, Eds.). Erlbaum Associates, Hillsdale, NJ.